Test and Evaluation in the "New World"

I wish to express my thanks to Gen. Farrell and the leadership of NDIA for affording me the opportunity to discuss with you some of my views and concerns with the new world of defense planning and defense acquisition. We have heard already about some of the activities underway or planned to confront the threat of terrorism and to transform our military forces and capabilities. My remarks today will focus on the new world of acquisition planning and strategies with some thoughts on how we, the test and evaluation community, can best serve our customers in this changing environment.

But first, permit me to pose the question of what is this new world in Defense Acquisition and Defense Planning? Is it Capabilities-Based Force Planning? Is it Commercial-off-the-Shelf procurement? Is it the use of Non-Development Items? Is it Simulation-Based Acquisition? Is it Commercial Best Practices? Is it Cost As an Independent Variable (CAIV)? Is it Total System Performance Responsibility? Is it Evolutionary Requirements? Is it Spiral Development?

I hate to sound like a "stick-in-the-mud" or some sort of Cassandra and I hate to admit that perhaps I've been around too long, but I can't help but say that I've seen and heard all this before. I've seen and participated in numerous Defense and Acquisition Reform exercises, task forces, working groups, IPTs -- you name it -- over the past 30 years or so. To use an overused cliché, I've been there, done that -- time and time again.

Early Reform Efforts

Let me spend a few minutes on a few of those efforts and their attempts to get at the problems besetting DoD's acquisition process. Not surprisingly, my remarks and commentary will emphasize the T&E aspects. Of particular note is the fact that the issues faced by these task forces throughout 30 years have been essentially the same -- to whit, how to get more effective and suitable systems into the hands of our operational combat forces more rapidly and at less costs to the taxpayer.

One of the most thorough and influential of these groups was the 1970 Blue Ribbon Defense Panel, also known as the Fitzhugh Commission, that addressed a whole host of defense management issues, to include *"Defense acquisition policies and practices, particularly as they relate to costs, time and quality."*

This Commission found the acquisition strategies in being then to be *"highly inflexible ... and also based on the false premise that technological difficulties can be foreseen prior to the detailed engineering effort on specific hardware."* The Commission went on to call for a major revision in policy with emphasis on several points, to include:

- Flexibility in selecting the techniques best suited to each new system.

- Proof of technology by building and testing hardware, including system prototypes where appropriate.

- Incremental development of subsystems and components separate from the development of major weapon systems.

- More gradual program build-ups.

This same Commission also made sweeping recommendations concerning the test and evaluation business, especially with respect to operational testing. The scene had been set for these recommendations by an earlier President's Scientific Advisory Committee report. That report stated, *"We regard the creation of the testing and evaluation group as of the utmost importance, since we believe most of our previous failures to be prepared for wars we fight would have been thoroughly exposed had an adequate program of testing and evaluation existed. The actual tests are very expensive and since the Testing and Evaluation budget in a Service is often in competition with funds for new equipment developments, we believe it is vital that the Test and Evaluation group in OSD have a substantial budget to allocate for tests."*

With respect to OT&E, the Blue Ribbon Presidential Commission made several cogent observations. Let me recall four of them:

- *It has been customary to think of OT&E in terms of physical testing. While operational testing is a very important activity ... it is emphasized that the goal is operational evaluation and that physical testing is only one means of attaining that goal. This is an important point since it is often argued that operational testing must await production of an adequate number of operationally-configured systems; and, by this time, it is too late to use the information gathered to help decide whether to procure the new system or even influence in any significant way the nature of the system procured.*

- *If OT&E, as a total process, is to be effective, it must extend over the entire life cycle of a system, from initial requirements to extending its life by adaptation to new uses. It must use analytical studies, operations research, systems analysis, component testing, testing of other systems, and eventually testing of the system itself.*

- *There is no effective method for conducting OT&E that cuts across Service lines although, in most actual combat environments, the U.S. must conduct combined operations.*

- *Because funds earmarked for OT&E do not have separate status in the budget, or in program elements, they are often vulnerable to diversion to other purposes.*

The 1986 President's Blue Ribbon Commission on Defense Management -- also known as the Packard Commission -- recommended a series of sweeping changes to the overall defense organization, its planning and budgeting as well as its acquisition processes,

many of which were implemented in statute with the Goldwater-Nichols Act later that year.

We recall two of the more than a dozen recommendations made by the Packard Commission with respect to the Acquisition process:

- *A high priority should be given to building and testing prototype systems and subsystems before proceeding with full-scale development. This early phase of R&D should employ ... streamlined procurement practices ... and should demonstrate that the new technology under test can substantially improve military capability, and should as well provide a basis for making realistic cost estimates prior to a full-scale development decision. This increased emphasis on prototyping should allow us to "fly and know how much it will cost before we buy."*

- *The proper use of operational testing is critical to improving the operations performance of new weapons. We recommend that operational testing begin early in advanced development and continue through full-scale development, using prototype hardware.*

At the risk of getting overly repetitive and, for that matter, restating the obvious, let me quote from one more of these defense management studies -- the Defense Management Report to the President in 1989 -- that was conducted pursuant to President Bush's direction in National Security Review #11.

"Decisions made during the early phases of systems development -- including those that involve funds and schedules for prototyping and testing -- often have dramatic consequences for operational performance and life-cycle cost. The Undersecretary of Defense for Acquisition will be charged with developing and ensuring rigorous application of policies that support sound decisions on major programs ...these policies will dictate that the schedules and management plans for major programs:

- *Support the building and testing of system and critical subsystem prototypes, the use of system engineering, and the validation of manufacturing processes as early as possible and certainly well prior to the commencement of high rate production;*

- *And provide for early test and evaluation of prototype hardware to prove concept, performance and suitability in realistic operational environments".*

Reform in the 1990s

Then, as we entered the 1990s, reform efforts -- for the most part based on internal reviews -- focused on specific aspects of the process that seemed to be at the core of the stretch-outs and cost growth that continued to plague our acquisition programs. One of the early initiatives called for establishing and adhering to exit criteria for future milestone decisions as a means of introducing much-needed discipline in the process. In

addition, the allegedly archaic requirements process was highlighted as the root cause for gold plating that was leading to more lengthy and expensive system developments. Several initiatives evolved from this latter determination -- among them Cost As an Independent Variable, Evolutionary Requirements, Capabilities-Based Acquisition, and even Spiral Development.

While these efforts have not lost sight of the ultimate objective of the acquisition process -- to provide the best and most capable weapon systems to our operational forces -- a prevalent, if not overarching, theme in many of them continues to be the desire, indeed the need, to cut the overall cycle time substantially. In many cases, their charters and objectives have been quite explicit: to cut the development and test time in half. Time is money and half the time required for development and test obviously should yield significant savings, get new technologies and capabilities to the war fighter before they become obsolete, and permit the purchase and fielding of larger quantities of more modern systems.

How Are We Doing?

But where are the results after decades of such initiatives? Have they not worked? Why is it that every three or four years or so, we seem to have this compelling need to reform the acquisition process once again, to formulate new strategies, to rewrite the 5000 series of directives and instructions, and to seek innovative approaches for reducing the time and effort required to field new military systems and capabilities? With all of these initiatives, Presidential Commissions, Defense Science Board Task Forces, etc., why do we still have major acquisition programs taking fifteen to twenty years -- and costing an arm and a leg -- to get through the process?

Just look at the track record of some of our current, high-visibility acquisition programs. Once again, showing my age and no doubt supporting the contention that I've been around too long, I recall the birth of the Comanche program in 1981 when the Army POM showed up with the new LHX program planned to replace the Army's fleet of UH-1 utility and AH-1 Cobra Attack helicopters. Twenty-one years and over five billion dollars later, what do we have to show for it? And we're still seven or more years from fielding the Comanche.

I attended the Milestone 0 review for the joint Army/Marine Corps JVX program in the fall of 1981, its Milestone I in 1982 and later the Milestone II review in 1986 that approved the program's entry into Full Scale Development. Today, and again more than twenty years and at least eleven billion dollars later, we have a V-22 Osprey program that is still several years from deployment with our operational forces as a much needed -- and long overdue -- replacement for the Marine Corps fleet of CH-46s.

In a similar vein, I recall the debate in the early- to mid-1980s about the scope and requirements for the Air Force's Advanced Tactical Fighter program. I took part in the DAB review in the fall of 1986 that approved entry into the Demonstration/Validation (Dem/Val) phase for this program. More than fifteen years and twenty-seven billion

dollars later, we're still at least a year away from IOT&E and several years from Full Operational Capability for the F-22.

Clearly, we did not plan nor foresee such long and drawn out schedules for these programs. We are past -- in fact, well past in some cases-- their originally planned IOCs. While I've obviously singled out three high visibility programs -- all aircraft developments undertaken in the 1980s -- I dare say that you will find similar histories for a whole host of our cutting edge development programs over the past twenty or so years.

So what, may I ask, has gone wrong? I recall being involved with a Defense Science Board panel in 1990 that was tasked with finding the root causes of the program stretch-outs and cost increases rampant during the late 1980s. We found that the dominant cause to be underestimates, or lack of appreciation, of the technical challenges faced by the programs at their outsets. Program after program had entered full-scale development before they were ready based on an overly optimistic view of the technical difficulties that lay ahead. When these technical problems surfaced, primarily in development testing, the stretch-out and cost growth cycle began. We found that the messenger of bad tidings, most often the T&E community, was then tarred with responsibility for the stretch outs.

My take on the history of all these reform efforts and their alleged failures is that we've really never taken them to heart; granted, we have made, and continue to make, changes to the process, but I feel we have yet to really come to grips with some of the root causes.

Clearly, we have failed time, and again, to do our homework early-on or to make the up-front investments required for an informed understanding of the technical and cost risks inherent in a program before we launched off into full-scale development and procurement. Too often, over the past few years, we've rushed into operational testing when the results of DT&E have clearly shown us that we were not ready and that our chances of success were minimal. In essence, we have been "rushing to failure."

I take exception to the views of those who continue to depict the T&E community as a major roadblock in the way of acquisition reform. The T&E community has done a truly commendable job, over the years, of responding and adjusting to accommodate these new and innovative approaches to acquisition. A good example is the initiative on the part of the T&E community, particularly the operational testers, to become involved in the early stages of acquisition developments -- an initiative espoused so effectively by my predecessor, Phil Coyle. This has proven tremendously beneficial to those programs that welcomed such involvement.

All that being said, the T&E community must continue to be flexible and ready to rapidly adjust to continuing efforts to streamline the acquisition process. In this new world of transformation, of rapid technology advances, and of less well-defined threats compared to the days of the Cold War, the acquisition process will no doubt continue to evolve to better serve our operating forces. So must we be willing to transform our policies and

practices where necessary to better serve those same forces as well as senior decision-makers.

First and foremost, and at the risk of stating the obvious, the T&E community must keep in mind that the principal reason for testing is to learn and to gain knowledge and information about the system being designed or in development. My predecessor put it so well in his oft-repeated dictum, "Testing is for learning!" That may sound somewhat trite, but how often have we strayed from that dictum and reflected the proverbial "Pass/Fail" mentality we're so often accused of.

Granted, there are certain statutory requirements that require evaluations of the adequacy of testing and the operational effectiveness and suitability of systems when measured against a set of predetermined standards or requirements before they enter full rate production. But, we as a community must be responsive, first and foremost, with the best information and knowledge available for customers that range from those involved in designing and developing weapons to those responsible for fielding and operating them in our combat forces.

Secretary Rumsfeld, speaking about 21st Century Transformation a month ago, stated:

"Preparing for the future will require us to think differently and develop the kinds of forces and capabilities that can adapt quickly to new challenges and to unexpected circumstances. An ability to adapt will be critical in a world where surprise and uncertainty are the defining characteristics of our new security environment."

We, the testing community, have a critical role to play in this new environment. While our record of adapting to new acquisition schemes in the past has indeed been exemplary, we can't afford to sit on our laurels. We must seriously consider restructuring ourselves, if necessary, as well as our thinking to better meet the new challenges referred to by the Secretary. Let me touch on just a couple of the new approaches to this changing environment that we should consider.

Dealing with Capabilities-Based Acquisition Programs

Among the major new initiatives is Capabilities-Based Acquisition. The idea here, as I see it, is a continuous process of design, development and testing of a new concept or system until we demonstrate and validate a level of capability deemed worth considering for procurement and deployment. At that point, the decision-maker -- hopefully, based on the informed advice of the potential user as well as the acquisition and testing communities -- decides that the system has indeed demonstrated a needed warfighting capability and approves advancing it, perhaps into full-scale engineering development, or even directly into production and deployment to our operational forces. One of the features of this approach is that, up to this point, there are no hard and fast requirements, threat-based or otherwise, against which to measure the operational effectiveness or suitability of the system.

As you no doubt know, this approach has been established as the acquisition strategy for the programs that fall under the newly established Missile Defense Agency; other programs are also considering it as a strategy.

How all this will work in detail is still a little murky. For example, I can imagine the difficulty for you in industry who might now work on big projects without a clear specification of what might be produced in the end. This could be quite challenging when you might not know until late in the game that production will be required of your system and what configuration that system will have. In essence, manufacturing planning may become a big problem, requiring innovative new thinking on your part.

For us in the testing community, one of the more obvious approaches to Capabilities-Based Acquisition is to move further away from the so-called pass/fail mentality to one of providing independent assessments of the capabilities (and limitations) of the system as tested to-date. In short, we will be called on for achievement-based reporting, so to speak, as opposed to activity-based reporting. We won't be making judgments as to effectiveness or suitability against some ORD-based bench-marks, but rather presenting our best judgment as to the capability demonstrated to-date in whatever environments -- open-air testing, hardware-in-the loop, or human-in-the-loop -- the system has been subjected to.

On the other hand, we may have to define, for our own use, some marks on the wall up-front with respect to capabilities. While the time at which those capabilities will be realized may not be known, when they, in fact, are achieved, a capable test infrastructure to include appropriate targets, instrumentation, and scenarios, will have to be available at our test ranges and facilities.

The T&E community must also view the operational forces that will deploy a new capability as a prime customer for its evaluations. In fact, I see a critical need to expand our contacts with operational users across-the-board and to cultivate them as principal recipients of our assessments -- and, I might add, not just in the context of this Capability-Based initiative.

Dealing with Spiral Development or Block Capability Programs

While Spiral Development and Block Upgrades might be somewhat different animals, their treatment by the T&E community is somewhat similar. As an aside, we have quite a bit of experience with such approaches, particularly in testing software-intensive systems to include the myriad of automated information systems. Here, we will plan our T&E strategies to assess incremental improvements in capabilities as opposed to using the full-up, or ultimate, system requirements spelled out in an operational requirements document as a bench-mark. At the least, our assessments will consider whether each spiral or block provides a measurable improvement in military capability over its predecessor.

Undoubtedly, the biggest financial commitment by a program in this context will be to

field the first spiral or Block I. Therefore, at a minimum, Block I should clearly demonstrate that it does not represent a decrease in military capability over legacy systems. In addition, if new functionality is added in a spiral or block, we will probably need to carry out some level of regression testing. There will also have to be some assessment of the growth potential of this spiral or block.

The new functionality -- if it is to be worth the disruption to the force by requiring retraining, additional training or new operational concepts -- ought to represent a significant improvement that should be easy to confirm. We should accept it as our responsibility to confirm, not only that improvement, but that the system continues to be effective and suitable for combat after fielding.

In spiral developments, we may need a formal feedback mechanism -- spiral reporting, so to speak -- to ensure that problems or deficiencies identified in T&E for each spiral are addressed and corrected by the program office. The information needs during spiral development seem to include at least: 1) what is the added capability of the new spiral, 2) what direction should the next spiral take to address the residual deficiencies of the incomplete system and 3) is the new spiral's increase in capability worth the disruption of introducing it into the force -- the reconfiguration, the revised training or the changed tactics, techniques and procedures the new spiral might imply.

Statistical significance of course may become less important in testing follow-on blocks or later spirals and it is certainly not what is most important to me, or the user. Instead, operational significance should be our focus. In general, the application of expert judgment about the changes in subsequent spirals or blocks coupled with operational testing can make it easier to confirm the system is effective and suitable with less testing, than if I knew nothing about the nature of the changes.

These considerations lead me to a need for some form of continuous testing, evaluation and reporting even after the system is deployed. Presumably, with increased use of spirals, there will be many more potential engineering change proposals. Hopefully, priorities accorded these proposals will be based on evaluation of data that shows what needs to be fixed depending on the most value to the war fighter.

Dealing with Cost As an Independent Variable or CAIV

As opposed to spiraling upward to greater and greater capability, CAIV has been employed as a way to spiral down to lower expectations in order to remain within some cost bogey. One idea that may prove useful in providing an evaluation of such a program is to plan the test against the original ORD and report the results against both the original ORD and the CAIVed-in ORD. We may have contracted to pay the developer for a system that meets the original ORD, but through the use of CAIV, we may have agreed to accept a capability short of those expectations. Implementing CAIV may prove to be a management problem, not just for contracting officers and lawyers, but a real problem for test managers. This is because the requirements might be changing faster than the Services can re-plan their tests.

New programs may be different. Certainly the original idea behind CAIV was not that the requirements would change just before the full-production decision. The original intent was to allow an informed discussion of the technical risks and costs in seeking specific capabilities before entering engineering and manufacturing development.

Summary

So, what's the bottom line after this tour-de-force of acquisition reform, or acquisition excellence in today's parlance, and some considerations for the T&E community in this new world?

First and foremost, we have a lot to be proud of over the past several years in our demonstrated flexibility and responsiveness to an ever-changing acquisition landscape. In fact, we've been involved in the T&E of spiral or block developments for quite some time now. We should step back now and translate our lessons learned in this context into more concrete policies or strategies for the future.

Our record of buying into early involvement and the fruits of that involvement are also praiseworthy. We have no choice but to continue and even expand our involvement earlier and continuously throughout the life cycle of systems. But, I am concerned with the increasing demands on our resources necessary to make those involvements continue to pay off.

We need to do much more in cultivating and serving the users, the operational forces and even the CINCs themselves as prime customers for our products. They need to know the capabilities and limitations of the new systems they are deploying, based on our best estimates of what the testing to-date has demonstrated. Where else, may I ask, will they get that information in a timely and straightforward fashion?

As I pointed out earlier, each evolution of acquisition reform has aimed at making substantial reductions in cycle-time, by at least a half in most cases. We, the testing community, should be looking at ways of cutting testing turn-around times in half. When the F-22 program fires but one missile a month in its test program, there is something profoundly wrong with the speed at which we can conduct testing.

Unfortunately, I am concerned that our T&E infrastructure is not in the best of shape needed to meet the challenges of the future. Failures of the acquisition process in the past, with all the program slips, have tended to ease the burden faced by the test ranges. Lord knows what would happen if all the programs that claimed to be ready for testing in 2002 actually showed up for testing. If the latest acquisition initiatives deliver what they hope for, then a greater fraction of programs should be ready for testing on or near their schedules. In this respect, I fear the T&E community might not be prepared for success in acquisition reform.

In closing, we, the T&E community -- in both industry and government, both technical and operational testers -- has served the department very well over the years. The success of our operational forces in the last several conflicts reflect that dedication to deploying systems proven effective and suitable on our ranges and in our facilities. But, there is a new world dawning that calls for new and innovative strategies and capabilities for T&E. I am confident that, together, we will rise to the challenge as we have in the past and ensure that our soldiers, sailors and airman are equipped with the best equipment our nation can provide.